(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
15 March 2001 (15.03.2001)

PCT

(10) International Publication Number
WO 01/18245 A2

(51) International Patent Classification[7]: C12Q 1/68

(21) International Application Number: PCT/IB00/01283

(22) International Filing Date:
6 September 2000 (06.09.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

09/391,244	7 September 1999 (07.09.1999)	US
09/443,539	19 November 1999 (19.11.1999)	US
09/470,673	23 December 1999 (23.12.1999)	US
09/471,598	23 December 1999 (23.12.1999)	US

(71) Applicant *(for all designated States except US)*: **DECODE GENETICS EHF.** [IS/IS]; Lynghals 1, IS-101 Reykjavik (IS).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **GRANT, Straun** [GB/IS]; Baronsstig 16, IS-101 Reykjavik (IS). **BLONDAL, Thorarinn** [IS/IS]; Lyngmoar 3, IS-210 Gardabaer (IS).

(74) Agent: **RAGNARSSON, Olafur**; Patice Iceland, Skeifan 7, P.O. Box 8630, IS-128 Reykjavik (IS).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *Without international search report and to be republished upon receipt of that report.*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/18245 A2

(54) Title: DETECTION OF ALTERATIONS IN A GENE BY LONG RANGE PCR USING HUMAN MOBILE ELEMENTS

(57) Abstract: Methods are described for detecting an alteration in a gene of interest, such as a deletion in the C4A gene (e.g., for detecting C4AQ0), by performing long range polymerase chain reaction amplification on a test sample comprising genomic DNA. The methods amplify target DNA comprising all or a portion of a human mobile element (e.g., a retroviral insert in intron 9 of the C4A gene), using primers designed such that PCR products are formed only if the test sample comprises genomic DNA comprising the alteration in the gene of interest; alternatively, the methods amplify target DNA using primers designed such that PCR products are formed only if the test sample comprises genomic DNA that does not comprise the alteration in the gene of interest. Alternatively, primers are designed such that PCR products have detectably different sizes, depending on whether or the test sample comprises genomic DNA that comprises the alteration in the gene of interest. The methods can be used to identify whether an individual is at risk for a disease or condition associated with the alteration in the gene of interest, as the presence of the alteration correlates with risk for the disease.

DETECTION OF ALTERATIONS IN A GENE BY LONG RANGE PCR USING HUMAN MOBILE ELEMENTS

RELATED APPLICATIONS

This application is a Continuation-in-Part of U.S. Serial No.: 09/471,598, filed on December 23, 1999, which is a Continuation-in-Part of U.S. Serial No.: 09/470, 673, filed on December 23, 1999, which is a Continuation-in-Part of U.S. Serial No.: 09/443,539, filed on November 19, 1999, which is a Continuation-in-Part of U.S. Serial No.: 09/391,244, filed September 7, 1999, the entire teachings of which are incorporated herein by reference.

BACKGROUND OF THE INVENTION

Systemic lupus erythematosus (SLE) is an autoimmune disease characterized by immune dysregulation resulting in the production of anti-nuclear antibodies, the generation of circulating immune complexes, and the activation of the complement system. SLE leads to inflammation of various parts of the body, especially the skin, joints, blood, kidneys, lungs, heart and nervous system. SLE affects approximately 1 in every 500 Americans, and strikes women 10-15 times more frequently than men. It is more common among Asians, and in China, SLE may be even more common than rheumatoid arthritis.

Although there is evidence of genetic etiology, linkage analysis suggests that there are no 'major' susceptibility genes segregating in families with SLE (Shai, R., *et al., Hum. Mol. Genet. 8*:639-644 (1999)). Nevertheless, a number of studies have demonstrated an association between SLE and certain major histocompatibility complex (MHC) antigens (Fielder, A.H. *et al., Br. Med J. (Clin. Res. Ed.)286(6363)*:425-8 (1983); Christiansen, F.T. *et al., Aust. NZ J. Med. 13(5)*:483-8 (1983); Reveille, J.D. *et al., Immunogenetics 21(4)*:299-311 (1985); Howard, P.F. *et al., Am. J. Med. 81(2)*:187-193 (1986); Kemp, M.E. *et al., Arthritis Rheum.*

30(9):1015-1022 (1987)). Within the MHC region, found on the short arm of chromosome 6, are the genes for the fourth component of the complement system: C4A and C4B. These genes are highly homologous, with only an 8 nucleotide difference in exon 26. This difference leads to six amino acid changes which allow resolution on sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) (Belt, K.T., *et al., Cell 36(4)*:907-914 (1984)).

HLA-DR3 and a C4A null allele are frequently co-inherited as the extended haplotype B8,BfS:C2C,C4AQ0,C4B1;DR3. This is the most common extended haplotype in white SLE patients (Kemp, M.E. *et al., Arthritis and Rheumatism 30*:1015-22 (1987)). Several C4AQ0-containing haplotypes have a DNA deletion of approximately 30 kB, extending from the 5' end of the C4A gene to the same position in the C4B gene. This deletion has been classically identified using Southern blotting, and has been found to be a genetic marker for SLE (Kemp, M.E. *et al., Arthritis and Rheumatism 30*:1015-22 (1987)). Southern blotting, however, is a time consuming and labor intensive process.

SUMMARY OF THE INVENTION

The invention pertains to methods of identifying an alteration in a gene of interest, particularly a gene of interest in the major histocompatibility region, utilizing long range polymerase chain reaction (LR-PCR) amplification of target DNA that includes all or a portion of a human mobile element. The alteration can be, for example, a deletion, insertion, duplication, or inversion. The human mobile element can be a DNA-based transposable element (e.g., a mariner element), autonomous retrotransposon (e.g., a retrotransposon containing long terminal repeats (LTRs), such as human endogenous retroviruses (HERVs); or lacking LTRs, such as L1 elements), or a non-autonomous retrotransposon (e.g., Alu element, pseudogene).

For example, a deletion in the C4A gene which extends from the 5' end of the C4A gene to the same position in the C4B gene, and which serves as a marker for systemic lupus erythematosus (SLE), can be detected. The deletion, referred to herein as "C4AQ0," is an approximately 30 kb deletion that is associated with the extended haplotype B8,BfS:C2C,C4AQ0,C4B1;DR3, which is the most common extended

haplotype in white SLE patients (Kemp, M.E. *et al., Arthritis and Rheumatism 30*:1015-22 (1987)). The methods can be used to determine whether an individual is at risk for developing systemic lupus erythematosus, as the presence of C4AQ0 correlates with a risk of developing systemic lupus erythematosus. The methods can additionally be used to determine the C4A deletion genotype of an individual (e.g., whether an individual is homozygous for C4AQ0; heterozygous for C4AQ0; or homozygous for the absence of C4AQ0).

In one embodiment, a test sample of genomic DNA is subjected to long range polymerase chain reaction (LR-PCR) amplification of target DNA that includes a human mobile element (e.g., an endogenous retroviral insert in intron 9 of the C4A gene); the LR-PCR primers are designed such that if the genomic DNA comprises an alteration (e.g., a deletion in the C4A gene), PCR products are formed, and if the genomic DNA does not comprise an alteration (e.g., no deletion in the C4A gene), no PCR products are formed. The presence or absence of PCR products is indicative of the presence or absence of the alteration (e.g., the presence or absence of the deletion in the C4A gene). In a preferred embodiment, the gene of interest is the C4A gene; the alteration is a deletion; and primers include a forward primer that corresponds to a DNA sequence in the G11 gene (e.g., TCTAGCTTCAGTACTTCCAGCCTGT (SEQ ID NO:1)), and a reverse primer corresponding to a DNA sequence in exon 10 of the C4A gene (e.g., GATGACACAAAATACCAGGATGTGA (SEQ ID NO:2)). These primers yield PCR products of approximately 5.4 kb in the presence of C4AQ0, and yield no detectable PCR products in the absence of C4AQ0 (i.e., in the absence of the deletion in the C4A gene).

In a second embodiment of the invention, a test sample of genomic DNA is subjected to long range polymerase chain reaction amplification of target DNA that includes a human mobile element (e.g., including the junction between intron 9 and the endogenous retroviral insert in intron 9 of the C4A gene), using primers that are designed such that if the genomic DNA comprises an alteration (e.g., a deletion in the C4A gene), no PCR products are formed, and if the genomic DNA does not comprise an alteration (e.g., no deletion in the C4A gene), PCR products are formed. The presence or absence of PCR products is indicative of the absence or presence of the

alteration. In a preferred embodiment, the gene of interest is the C4A gene; the alteration is a deletion; and primers include a forward primer that corresponds to a DNA sequence in the G11 gene (e.g., TCTAGCTTCAGTACTTCCAGCCTGT (SEQ ID NO:1)), and a reverse primer corresponding to the junction between intron 9 and the retroviral insert in intron 9 of the C4A gene (e.g., TGGTCCCCAACATGTCTGTGCATGCTG (SEQ ID NO:3)); these primers yield PCR products of approximately 5.2 kb in the absence of C4AQ0 (i.e., in the absence of the deletion in the C4A gene), and yield no detectable PCR products in the presence of C4AQ0.

Alternatively, in yet another embodiment of the invention, the primers for the long range polymerase chain reaction amplification are designed such that PCR products having detectably different sizes are produced in the presence and in the absence of the alteration (e.g., in the presence or absence of C4AQ0); an assessment of the size of the PCR products indicates whether the alteration is present or absent.

The methods of the invention are simple to perform, provide consistent results, and can be adapted for high-throughput screening of test samples. In addition, the methods facilitate genotyping of individuals, thereby affording a quick and reliable means for identification of individuals at risk for inheriting or for developing diseases associated with alterations in a gene of interest, such as systemic lupus erythematosus.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a representation of long range polymerase chain reaction (PCR) detection of the presence of C4AQ0 (presence of the deletion in the C4A gene).

Figure 2 is a representation of long range polymerase chain reaction (PCR) detection of the absence of C4AQ0 (absence of the deletion in the C4A gene).

DETAILED DESCRIPTION OF THE INVENTION

The present invention relates to methods for detecting an alteration in a gene of interest. A gene "of interest," as used herein, refers to a gene (e.g., a nucleic acid encoding a polypeptide) which is to be assessed for the presence or absence of an

alteration. In a preferred embodiment, the gene is a gene in the major histocompatibility region (MHC). The methods can be employed, for example, in assessment of alterations associated with complement-based diseases, such as rheumatoid arthritis, ankylosing spondylitis, scleroderma, subacute sclerosing panencephalitis, systemic lupus erythematosus, or renal disease such as nephropathy (e.g., IgA nephropathy). An "alteration" in the gene is a change in the nucleic acid sequence of the gene (e.g., in a test sample) as compared to a known or expected nucleic acid sequence of the gene. The alteration can be, for example, a deletion (e.g., of one or more nucleotides), insertion (e.g., of one or more nucleotides), duplication (e.g., of all or a portion of the gene), or inversion.

In the methods of the invention, polymerase chain reaction (PCR) amplification of long DNA sequences (herein referred to as "long range PCR" or "LR PCR") is used to amplify DNA adjacent to and including a human mobile element (for a discussion of human mobile elements, see, for example, Kazazian, H. and Moran, J., *Nature Genetics* 19:19-24 (1998); the entire teachings of this reference is incorporated herein by reference in its entirety). The human mobile element can be a DNA-based transposable element (e.g., a mariner element). Alternatively, the human mobile element can be an autonomous retrotransposon, for example, a retrotransposon containing long terminal repeats (LTRs), such as short or long interspersed elements (SINEs or LINEs), or human endogenous retroviruses (HERVs), or a retrotransposon lacking LTRs, such as L1 elements. The human mobile element can also be a non-autonomous retrotransposon (e.g., Alu element, pseudogene). In a preferred embodiment, the human mobile element of interest is a human endogenous retrovirus (HERV). HERVs are particularly useful for analysis of genes of interest of the major histocompatibility complex (MHC), because of the prevalence of HERVs within the MHC, as well as the possible relationship between HERVs and complement-based or other MHC-related diseases (see, e.g., Andersson, G. *et al., Trends Genet.* 14(3):109-114 (1998); Kulski, J.K. *et al., J. Mol. Evol.* 48(6):675-83 (1999); Taruscio, D., and Mantovani, A., *Teratology* 53(2):108-110 (1996); Hasuike, S. *et al., J. Hum. Genet.* 44(4):343-7 (1999); Kulski, J.K. *et al., J. Mol. Evol.* 49(1):84-97 (1999)).

The human mobile element used in the methods is referred to as the "human mobile element of interest." The primers used in the LR PCR reaction are designed such that in the presence of an alteration in the gene of interest, a PCR product is produced which can be detected, and in the absence of the alteration, no PCR product is produced which can be detected, or a PCR product of a detectably different size can be detected. The absence of the alteration can be confirmed by LR PCR using primers designed such that in the absence of the alteration, a PCR product is produced which can be detected, and in the presence of the alteration, no PCR product is produced which can be detected (or PCR products of a detectably different size can be detected).

For example, in a particular embodiment of the invention, a deletion in the C4A gene is detected. In particular, the deletion C4AQ0, which extends from the 5' end of the C4A gene to the same position in the C4B gene, is detected. This deletion serves as a marker for systemic lupus erythematosus (SLE) (see, e.g., Kemp, M.E. *et al., Arthritis and Rheumatism 30*:1015-22 (1987); Arnett, F.C., *Clin. Immunol. Immunopathol. 63(1)*:4-6 (1992)). Methods for detecting a deletion in the C4A gene take advantage of a polymorphism in the C4A gene which results from the presence of a 6.4 kB retroviral insertion in intron 9 of the C4 gene (Chu, X., *et al., Exp. Clin. Immunogenet. 12*:74-81 (1995)). The primers used in the LR PCR reaction are designed such that in the presence of a deletion in the C4A gene (e.g., in the presence of C4AQ0), a PCR product is produced which can be detected, and in the absence of the deletion (e.g., in the absence of C4AQ0), no PCR product is produced which can be detected, or a PCR product of a detectably different size can be detected. The absence of the deletion can be confirmed by LR PCR using primers designed such that in the absence of the deletion, a PCR product is produced which can be detected, and in the presence of the deletion, no PCR product is produced which can be detected (or PCR products of a detectably different size can be detected). The methods can be used to identify individuals at risk for developing SLE or for inheriting SLE, or to confirm a diagnosis of SLE in an individual suspected of having SLE.

POLYMERASE CHAIN REACTION AMPLIFICATION

Polymerase chain reaction (PCR) amplification is a well-known tool for amplification of nucleotide sequences. During a single cycle of PCR amplification, a double-stranded target DNA sequence is denatured; primers are annealed to each strand of the denatured target; and the primers are extended by a DNA polymerase. This cycle is repeated, generally between 25 and 40 times, in order to concentrate the number of copies of a target DNA sequence in a sample. The primers used in PCR are designed to anneal to the denatured target DNA sequence strands in a position and orientation such that the extended primers are complementary copies of the target DNA sequences. On subsequent amplification cycles, the extended primers can also serve as targets for amplification. PCR is described in detail in U.S. Patent No.s 4,683,202; 4,683,195; 4,800,159; and 4,965,188; the entire teachings of these patents are incorporated herein by reference. "Long range" PCR utilizes amplification conditions which improve target strand denaturation (e.g., higher denaturation temperatures, addition of cosolvents), and which protect DNA from degradation; utilizes longer extension times; and minimize incorporation of erroneous nucleotides by utilizing polymerases having exonuclease activity to reduce mismatches, thereby enabling amplification of extended strands of DNA. Long range PCR is described in detail, for example, in U.S. Patent 5,512,462; in Burland, V. and Kusukawa, N., *Biotechniques 23*:1070-1072, 1074-1075 (1997)); and Cheng, S. *et al., Proc. Natl. Acad. Sci. USA 91*:5695-5699 (1994)); the entire teachings of which are incorporated by reference herein.

TEST SAMPLE

In the methods of the invention, a test sample comprising genomic DNA is used. The test sample is obtained from an individual suspected of having (or of carrying a defect associated with) an alteration in a gene of interest (e.g., an individual suspected of having or of carrying a defect associated with systemic lupus erythematosus (SLE)) (the "test individual"). The individual can be an adult, child or fetus. The test sample can be from any source which contains genomic DNA, such as a blood or tissue sample (e.g.. from skin or other organs). In a preferred embodiment,

the test sample is obtained from a blood sample, a fibroblast skin sample, from hair roots, or from cells obtained from the oral cavity (e.g., via mouthwash). In another preferred embodiment, the test sample is obtained from fetal cells or tissue by appropriate methods, such as by amniocentesis or chorionic villus sampling. The test sample is subjected to LR PCR amplification, and the presence or absence of an alteration (e.g., the presence or absence of a deletion in the C4A gene) is then detected.

PRIMERS

To conduct LR PCR, primers are designed to amplify DNA adjacent to, and/or including at least part of, a gene of interest. The DNA that is amplified contains all or a portion of a human mobile element of interest (e.g., the retroviral insert in the C4A gene), as well as the part of the gene (or DNA adjacent to the gene) which may contain the alteration of interest. In some embodiments, the human mobile element of interest can be a part of the gene of interest (e.g., not solely adjacent to the gene of interest), and can contain the alteration of interest. The DNA that is targeted for amplification, which contains the human mobile element of interest, is referred to herein as the "target" DNA. The term "primer," as used herein, refers to an oligonucleotide that is capable of serving as an initiation point for nucleic acid synthesis during PCR, under appropriate conditions as described below. The primer typically ranges from 15 to 50 nucleotides , and/or has a T_m of approximately 50-75°C; preferably, the primer is approximately 25-30 nucleotides in length, and/or has a T_m of approximately 60-65°C. Primers can be prepared by a variety of methods, including chemical synthesis (see, e.g., Narang *et al., Meth. Enzymol. 68*:90-99 (1979); Brown *et al., Meth. Enzymol. 68*:109-151 (1979); Beucage *et al., Tetrahedron Lett. 22*:1859-1862 (1981); U.S.patent 4,458,066; the entire teachings of these references are incorporated herein in their entirety). The primers are designed such that the PCR products (as described below) obtained from the primers will differ in size, depending on the presence or absence of alteration (e.g., the presence or absence of a deletion in the C4A gene, such as C4AQ0). That is, in the presence of the

alteration, the primers will yield PCR products of a certain (first) size, and in the absence of the alteration, the same primers will yield PCR products of a (second) size that is detectably different from the size of the PCR products in the presence of the deletion (the first size). A "detectably different" size indicates that the differences in the sizes of the products can be identified, using standard techniques, such as described below. In another embodiment, the primers are designed such that no PCR products are produced either in the presence or in the absence of the alteration. For example, in one embodiment, the primers will yield PCR products of a certain size in the presence of the alteration, and will yield no PCR products in the absence of the deletion. Alternatively, in another embodiment, the primers will yield PCR products of a certain size in the absence of the alteration, and will yield no PCR products in the presence of the alteration.

The nucleotide sequences of the primers correspond to DNA sequences adjacent to or present in the human mobile element of interest (e.g., adjacent to the C4A gene or the retroviral insert). A primer that "corresponds to" a DNA sequence is a primer that has the same nucleotide sequence as the DNA sequence, or that is sufficiently complementary to the DNA sequence that it hybridizes under PCR conditions to the DNA sequence. A DNA sequence that is "adjacent to" the a human mobile element of interest is DNA sequence that is in physical proximity to the mobile element of interest. For example, a DNA sequence that is "adjacent to" the C4A gene is a DNA sequence that is in physical proximity to the C4A gene on chromosome 6, such as a DNA sequence in a gene next to the C4A gene (e.g., the G11, C2, TEN, Bf, 21-OH or RD genes, preferably the G11 gene). An "upstream" or "forward" primer (a primer that hybridizes to the non-coding strand of the target DNA and forms the 5' end of the amplified product of the coding strand) and a "downstream" or "reverse" primer (a primer that hybridizes to the coding strand of the target DNA and forms the 5' end of the amplified product of the non-coding strand) are used.

To detect the presence or absence of an alteration in a gene of interest, the forward primer corresponds to a unique DNA sequence upstream of the part of the gene of interest that may contain the alteration, and the reverse primer corresponds to

a unique DNA sequence downstream of the part of the gene of interest that may contain the alteration; either the forward primer, or the reverse primer, comprises all or a portion of the human mobile element. For example, to detect the presence of a deletion in the C4A gene (e.g., the presence of C4AQ0), the forward primer corresponds to a unique DNA sequence upstream of the target DNA (e.g., a forward primer corresponding to a DNA sequence in the G11 gene), and the reverse primer corresponds to a DNA sequence in an exon of the C4A gene that is after (downstream of) the location of the target DNA which is a retroviral insert (e.g., the second half of intron 9, exon 10, or beyond). To detect the absence of a deletion in the C4A gene (e.g., to detect the absence of C4AQ0, or to detect the presence of the retroviral insert), the same forward primer as described above is used; the reverse primer corresponds to a DNA sequence within the target DNA which is a retroviral sequence (e.g., the DNA sequence in the junction between the retroviral sequence and intron 9). In one embodiment of the invention, to detect the presence of C4AQ0, the forward primer TCTAGCTTCAGTACTTCCAGCCTGT (SEQ ID NO:1), which corresponds to a DNA sequence in the G11 gene, is used, and the reverse primer GATGACACAAAATACCAGGATGTGA (SEQ ID NO:2), which corresponds to a DNA sequence in exon 10 of the C4A gene, is used. In another embodiment of the invention, to detect the absence of C4AQ0, the same forward primer TCTAGCTTCAGTACTTCCAGCCTGT (SEQ ID NO:1) is used, and the reverse primer TGGTCCCCAACATGTCTGTGCATGCTG (SEQ ID NO:3), which corresponds to a DNA sequence of the junction between the retroviral sequence and intron 9 of the C4A gene, is used.

PCR CONDITIONS

The test sample of genomic DNA and the primers are used in LR PCR amplification. Long range PCR amplification is performed, for example, as described in U.S. Patent 5,512,462. Briefly, the test sample of genomic DNA and the primers are mixed in an amplification reaction mixture. An "amplification reaction mixture" contains reagents necessary for amplification of the target DNA sequence (e.g., nucleotides, enzymes, buffers, etc.). Representative amplification reaction mixtures

include commercial kits, such as the rTth DNA POLYMERASE XL kit (Perkin Elmer, catalog number N808-0187). The amplification reaction mixture, including the test sample and primers, is then subjected to cycles of varying temperature. If a commercial kit is used, the manufacturer's suggested protocol can be used for amplifying the target DNA.

For example, in one embodiment of the invention, a 20 µl reaction mixture can be used, including the following components:

Template DNA (genomic)	1 µl (approximately 100 ng)
3.3 x Buffer	6 µl (e.g., rTth Polymerase XL kit buffer, Perkin-Elmer)
MgOAc	0.8 µl (Final: 1 mM)
dNTP (2 mM)	2 µl (Final: 200 µM)
Forward primer (5 µM)	2 µl (Final: 0.5 µM)
Reverse primer (5 µM)	2 µl (Final: 0.5 µM)
rTth polymerase	0.4 µl**
Water	5.8 µl

** rTth pol has exonuclease activity; PCR is started immediately after addition of rTth pol, which is added last to the mixture.

The reaction mixture is then subjected to cycling conditions, such as the following:

1 cycle:	94°C for 30 sec
35 cycles:	90°C for 10 sec
	55°C for 10 sec
	65°C for 10 min.
hold at 4°C.	

ANALYSIS OF PCR PRODUCTS

Following the LR PCR amplification reaction, the PCR products, if any, are detected. The term, "PCR products," refers to copies of the target DNA sequence that are produced during PCR amplification (i.e., DNA which has been amplified during

the PCR process). If no DNA has been amplified during PCR, no PCR products will be generated. Analysis of the PCR products includes detecting the presence (or absence) of detectable PCR products; in a preferred embodiment, analysis of the PCR products includes determining the size of any detectable PCR products. The PCR products can be detected by a variety of methods; in a preferred embodiment, methods which separate the DNA by size, such as gel electrophoresis (e.g., agarose or acrylamide gel electrophoresis), or HPLC, are used to separate PCR products, and are followed by detection of the size fractionated DNA by methods such as staining (e.g., with ethidium bromide), or hybridization of labeled probes. Representative methods are described in *Current Protocols in Molecular Biology*, Ausubel, F. *et al.*, eds., John Wiley & Sons, including all supplements through 1999. In a preferred embodiment, ethidium bromide agarose gel electrophoresis is used, and the presence or absence of PCR products is then detected.

DETERMINATION OF THE PRESENCE OF AN ALTERATION

The presence of absence of an alteration in a gene of interest (e.g., a deletion in the C4A gene) can be determined based on the presence or absence, or the size, of the PCR products. For example, as shown in Figure 1, if the test sample comprises genomic DNA containing C4AQ0, then a forward primer in the G11 gene and a reverse primer in exon 10 (e.g., SEQ ID NO:1 and SEQ ID NO:2) will yield a PCR product of approximately 5.4 kb. If the test sample comprises genomic DNA that does not contain C4AQ0, the DNA between these primers will be approximately 11.8 kb, which is too big a product to be generated (that is, no PCR products will be generated). Thus, detection of a PCR product, and particularly of a PCR product of approximately 5.4 kb, is indicative of the presence of a deletion, and particularly of the presence of C4AQ0 which is associated with SLE. Lack of a PCR product is indicative of the absence of C4AQ0. If a PCR product is detected (i.e., if the test sample contains genomic DNA that comprises a deletion, such as C4AQ0), further experiments can also be performed to determine whether the test sample is homozygous or heterozygous for the deletion (e.g., heterozygous or homozygous for C4AQ0). For example, a forward primer in the G11 gene (e.g., SEQ ID NO:1) is

used, and a reverse primer which corresponds to a DNA sequence of the junction between the retroviral sequence and intron 9 of the C4A gene (e.g., SEQ ID NO:3) is used. As shown in Figure 2, if the test sample comprises genomic DNA that does not contain the deletion (i.e., is heterozygous for C4AQ0), these primers will yield a PCR product of approximately 5.2 kb. If the test sample comprises genomic DNA that does contain the deletion (i.e., is homozygous for C4AQ0), no PCR products will be generated. Thus, detection of a PCR product, and particularly of a PCR product of approximately 5.2 kb, is indicative of the absence of the deletion, and particularly the absence of C4AQ0. Lack of a PCR product is indicative of the presence of the deletion, particularly the presence of C4AQ0 which is associated with SLE.

DETERMINATION OF RISK OF AN INDIVIDUAL FOR DEVELOPING A DISEASE ASSOCIATED WITH THE ALTERATION IN THE GENE OF INTEREST

Using the methods described above, a test sample of genomic DNA from an individual, such as an individual suspected of being at risk for developing or inheriting a disease or condition associated with an alteration in a gene of interest (e.g., SLE), is analyzed for the presence or absence of the alteration in the gene of interest (e.g., the deletion in the C4A gene, such as an analysis for the presence or absence of C4AQ0). If a test sample from the individual comprises genomic DNA that contains the alteration (e.g., a deletion in the C4A gene, such as the presence of C4AQ0), the presence of the deletion therefore indicates that the individual is at risk for developing or inheriting the disease or condition associated with the alteration in the gene of interest.

In addition, using the methods described above, a test sample of genomic DNA from an individual can be analyzed homozygosity or heterozygosity of an alteration of interest. For example, a test sample of genomic DNA can be analyzed for the presence or absence of deletions in the C4A gene (e.g., for the presence or absence of C4AQ0), and the C4A deletion genotype (for example, whether the individual is homozygous for (the presence of) C4AQ0; heterozygous for C4AQ0; or homozygous for the absence of C4AQ0) can be determined. For example, a test

sample comprising genomic DNA from the individual can be subjected to long range polymerase chain reaction amplification of target DNA comprising a retroviral insert in intron 9 of the C4A gene, as described above. If the test sample comprises genomic DNA comprising a deletion in the C4A gene, PCR products are formed, and if the test sample does not comprise genomic DNA comprising a deletion in the C4A gene, no PCR products are formed; thus, the presence of PCR products indicates that the individual is either homozygous or heterozygous for a deletion in the C4A gene (e.g., C4AQ0), and the absence of PCR products indicates that the individual is homozygous for the absence of a deletion in the C4A gene (e.g., C4AQ0). Homozygosity or heterozygosity for the deletion can be determined by subjecting a test sample comprising genomic DNA from the individual to long range polymerase chain reaction amplification of target DNA comprising a junction between intron 9 and retroviral insert in intron 9 of the C4A gene, as described above. If the test sample comprises genomic DNA comprising a deletion in the C4A gene (e.g., C4AQ0), no PCR products are formed, and if the test sample does not comprise genomic DNA comprising a deletion in the C4A gene, PCR products are formed; the absence of PCR products indicates that the individual is homozygous for a deletion in the C4A gene (e.g., C4AQ0), and the presence of PCR products indicates that the individual is heterozygous for a deletion in the C4A gene (e.g., C4AQ0). This method provides an advantage over protein-based methods of analysis, as the C4A protein is produced in very low amounts, rendering it difficult to determine genotype by analysis of the amount of C4A protein.

Genotyping allows an assessment of an individual's risk for developing (or inheriting) SLE. Individuals with no deletion are at a reduced risk of developing SLE, compared with individuals with one or more deletions. At least one C4AQ0 (i.e., a heterozygous or homozygous genotype) has been identified in up to 50% or more patients with SLE, and homozygosity for C4AQ0 is reported in 11-13% of SLE patients, compared to 2-3% of controls (Schur, P.H., Dubois' Lupus Erythematosus 5th ed:254-261 (1977); Arnett, F.C., Moulds, J.M., *Clin. Exp.Rheumatol. 9*:289-296 (1991); Ratnoff, W.D., *Rheumatic. Dis. Clin. N. Am. 22*:75-94 (1996); Kumar, A. *et al., Clin. Immunol. Immunopathol. 60*:55-64 (1991); Fielder, A.H. *et al., Br. Med. J.*

(Clin. Res. Ed.) 286 (6363):425-8 (1983); Reveille, J.D. *et al., Immunogenetics 21(4)*:299-311 (1985)).

USES OF INVENTION

The methods of the invention, which have been exemplified by an analysis of the C4A gene for the presence of absence of a deletion (e.g., for the presence or absence of C4AQ0), are equally applicable to analysis of other genes for the presence of absence of other alterations (e.g., deletions, duplications, insertions, and inversions) in a gene of interest. For example, as indicated above, the methods are particularly useful for the identification of alterations (e.g., deletions) within the genes of the major histocompatibility complex (MHC) region, as they enable analysis of sizable regions of DNA such as that of the MHC. Thus, methods can be employed, for example, in assessment of alterations associated with complement-based diseases, such as rheumatoid arthritis, ankylosing spondylitis, scleroderma, subacute sclerosing panencephalitis, and renal disease such as nephropathy (e.g., IgA nephropathy).

To use the present methods for analysis of a region of DNA for the presence or absence of a particular alteration in a gene of interest, primers are designed to amplify regions of DNA including and/or surrounding the alteration. Of particular interest is the annealing temperature (T_m) of the primers: the annealing temperature should be designed to ensure success of the polymerase chain reaction under the conditions necessary for long range PCR amplification. For example, as described above, the T_m of the primers should be approximately 50-75°C, preferably approximately 60-65°C. The primers are designed such that the PCR products, as described above, that are obtained from the primers will differ in size, depending on the presence or absence of the alteration in the gene of interest. That is, in the presence of the alteration, the primers will yield PCR products of a certain (first) size, and in the absence of the alteration, the same primers will yield PCR products of a (second) size that is detectably different from the size of the PCR products in the presence of the alteration (the first size). Alternatively, the primers are designed such that no PCR products are produced either in the presence or in the absence of the

alteration. For example, the primers will yield PCR products of a certain size in the presence of the alteration, and will yield no PCR products in the absence of the alteration; alternatively, the primers will yield PCR products of a certain size in the absence of the alteration, and will yield no PCR products in the presence of the alteration. A test sample is obtained as described above, and long range PCR is performed as described. The PCR products, if any, are then assessed for the presence or absence of an alteration in the gene of interest, using the methods as described above. The presence or absence of the PCR products, or the presence of PCR products of particular sizes, is indicative of the presence or absence of the alteration in the target gene of interest.

The following Exemplification is offered for the purpose of illustrating the present invention and is not to be construed to limit the scope of this invention. The teachings of all references cited herein are hereby incorporated herein by reference.

EXEMPLIFICATION Detection of the Presence or Absence of a Deletion in the C4A Gene that is Associated with SLE

C4 Protein Allotyping

C4 protein allotypes were analyzed by high-voltage agarose electrophoresis on carboxypeptidase (Sigma Type I) and neuraminidase (Sigma Type VIII) treated serum samples followed by immunofixation with monoclonal antibodies (Incstar) (sim and Cross, *Biochem. J. 239*:763-767 (1986)). C4A and C4B allotypes were determined by their positions on the gel in comparison to known standard samples. C4A and C4B zygosity was determined by their relative band intensity.

HLA Typing

Typing of MHC class allotypes (HLA-A, B, and C) was performed by serological methods using the lymphocytotoxicity test (Perdue *et al., Tissue Antigens 9*:259-266 (1977)). Typing of MHC class II alleles (HLA-DR and DQ) was performed by PCR with sequence specific primers (PCR-SSP) (Dynal) (Olerup and Zetterquist, *Tissue Antigens* 30:225-235 (1992)). A segregation analysis was done

using pedigree information to give haplotypes for MHC classes I, II and III. For individuals with incomplete typing results, the most probable haplotypes were deduced from first degree relatives.

Sample Selection

C4 protein allotyping and HLA typing was performed on a collection of DNA samples obtained in rheumatological disease studies performed at the Center for Rheumatology Research, Reykjavik, Iceland. This identified appropriate individuals for screening with a long PCR strategy. There were two individuals who were both C4A protein deficient and homozygous for B8-C4AQ0-C4B1-DR3, whilst 21 heterozygotes had protein deficiency of C4A. A total of 66 individuals did not have the B8-C4AQ0-C4B1-DR3 extended haplotype and were not deficient in C4A protein expression.

C4A Deletion Genotyping by Long PCR

Two separate long PCR assays - one, C4A deletion-specific and the other, non-deletion-specific - were designed to determine the C4A genotype. Individuals having different genotypes (homozygous for C4AQ0 (the "deletion"); heterozygous for the deletion; homozygous for no deletion) were assessed to determine the presence or absence of C4AQ0 using long range PCR. Genotypes of the individuals had been previously determined using analysis of C4 protein and HLA typing as described above.

Primers were designed based on the relative location of the G11 gene to intron 9 of C4A. The same forward primer in the G11 gene was used for both the C4A deletion and non-deletion detection. To detect the C4A deletion, a reverse primer specific for exon 10 of C4A/C4B was used. When the C4A deletion occurred, a 5.4 kb (5401 bp) product was expected to be generated, while in the absence of the C4A deletion, the potential 11.8 kb (11,775 bp) product was too large to be generated with the cycling conditions, due to the presence of the 6/4 kb (6374 bp) HERV-K(C4) in intron 9. In the case of a heterozygote for the deletion, long PCR would favor the

amplification of the 5.4 kb product over the 11.8 kb product. Thus, separate reactions specific for each circumstance were performed. To detect the absence of the C4A deletion, the reverse primer was designed to anneal at the junction between the HERV-K(C4) sequence and the unique sequence of intron 9 because the C4A deletion lacks this HERV-K(C4) sequence. When the non-deletion occurred (i.e., in the absence of the C4A deletion), a 5.2 kb (5237 bp) product was expected to be generated.

DNA from peripheral white blood cells was used in the samples. The samples were subjected to LR PCR amplification, using rTth Polymerase XL kit (Perkin-Elmer). The amplification reaction mixture included the following components, for a total volume of 20 µl:

Template DNA (genomic)	1 µl (approximately 100 ng)
3.3 x Buffer*	6 µl
MgOAc	0.8 µl (Final: 1 mM)
dNTP (2 mM)	2 µl (Final: 200 µM)
Forward primer (5 µM)	2 µl (Final: 0.4 µM)
Reverse primer (5 µM)	2 µl (Final: 0.4 µM)
rTth polymerase	0.4 µl
Water	5.8 µl

* Buffer containing tricine, potassium acetate, glycerol and DMSO.

For the forward primer, TCTAGCTTCAGTACTTCCAGCCTGT (SEQ ID NO:1) ("primer 1"), which corresponds to a DNA sequence in the G11 gene, was used. For the reverse primer, either GATGACACAAAATACCAGGATGTGA (SEQ ID NO:2) ("primer 2"), which corresponds to a DNA sequence in exon 10 of the C4A gene, or TGGTCCCCAACATGTCTGTGCATGCTG (SEQ ID NO:3) ("primer 3"), which corresponds to a DNA sequence in the junction between the retroviral sequence and intron 9 of the C4A gene, was used. Because rTth pol has exonuclease activity; PCR was started immediately after addition of rTth pol, which is added last to the mixture. The reaction mixture was subjected to the following cycling conditions:

1 cycle:	94°C for 30 sec
35 cycles:	90°C for 10 sec
	55°C for 10 sec
	65°C for 10 min.

hold at 4°C.

For electrophoresis, the PCR products were loaded with standard loading buffer (0.25% bromophenol blue; 40% (w/v) sucrose in water) on a 0.8% ethidium bromide agarose gel (Sigma), run in Tris-borate-EDTA (1xTBE) at 100 V/hr, and visualized under ultraviolet light (Eagle Eye II, Stratagene).

Cloning and End-sequencing of the Long PCR Products

Long PCR products were cut out of an agarose gel and purified using the ultrafree DA purification kit (Millipore) followed by ethanol precipitation. dA 3' overhang addition was done in 10 µl reactions in 200 µM dNTP, using 1 U Taq polymerase (Promega) and 1X Taq buffer. Reactions were carried out at 72°C for 15 minutes and terminated by cooling on ice. The long PCR product was cloned into the TOPO XL TA vector (Invitrogen) and transformed into TOP 10 cells according to the manufacturer. Positive clones were sequenced approximately 500 bp from each end to confirm accurate amplification using Universal M13 forward (-20) (16 mer) and reverse (17 mer) primers (Invitrogen) and the BigDye terminator cycle sequencing kit (Perkin Elmer). The samples were run on an ABI 377 and analyzed using sequence analysis softward 3.1 (Perkin Elmer).

Results

The results indicated that, for individuals with a homozygotic or heterozygotic deletion in the C4A gene extending from the 5' end of the C4A gene to the same position in the C4B gene (i.e., homozygotic or heterozygotic for C4AQ0), primers 1 and 2 yielded a 5.4 kB PCR product. For individuals with no deletion (i.e., homozygotic normal), no detectable PCR product was produced. For individuals with a heterozygotic deletion, primers 1 and 3 yielded a 5.2 kB PCR product.

Notably, the homozygotes for C4AQ0, and the heterozygotes for C4AQ0, all had SLE.

While this invention has been particularly shown and described with references to preferred embodiments thereof, it will be understood by those skilled in the art that various changes in form and details may be made therein without departing from the scope of the invention encompassed by the appended claims.

CLAIMS

What is claimed is:

1. A method of detecting the presence or absence of an alteration in a gene of interest, comprising:

   a) subjecting a test sample comprising genomic DNA to long range polymerase chain reaction amplification of target DNA comprising a human mobile element of interest, such that if the test sample comprises genomic DNA comprising the alteration in the gene of interest, PCR products are formed, and if the test sample does not comprise genomic DNA comprising the alteration in the gene of interest, no PCR products are formed; and

   b) detecting the presence or absence of the PCR products,

   wherein the presence of PCR products is indicative of the presence of the alteration in the gene of interest in the test sample, and the absence of PCR products is indicative of the absence of the alteration in the gene of interest in the test sample.

2. A method of detecting the presence or absence of an alteration in a gene of interest, comprising:

   a) subjecting a test sample comprising genomic DNA to long range polymerase chain reaction amplification of target DNA comprising a human mobile element of interest, such that if the test sample comprises genomic DNA comprising the alteration in the gene of interest, PCR products having a first size are formed, and if the test sample does not comprise genomic DNA comprising the alteration in the gene of interest, PCR products having a second size are formed, wherein the first size and second size are detectably different; and

   b) assessing the size of the PCR products,

wherein the presence of PCR products having the first size is indicative of the presence of the alteration in the gene of interest in the test sample, and the presence of PCR products having the second size is indicative of the absence of the alteration in the gene of interest in the test sample.

3. The method of Claim 1 or Claim 2, wherein the human mobile element of interest comprises the alteration in the gene of interest.

4. The method of Claim 1 or Claim 2, wherein the alteration is selected from the group consisting of: a deletion; a duplication; an insertion; and an inversion.

5. The method of Claim 1 or Claim 2, wherein the gene of interest is a gene of the major histocompatibility complex (MHC).

6. The method of Claim 1 or Claim 2, wherein the human mobile element of interest is selected from the group consisting of: a DNA-based transposable element; an autonomous retrotransposon; and a non-autonomous retrotransposon.

7. The method of Claim 6, wherein the autonomous retrotransposon is a human endogenous retrovirus or an L1 element.

8. The method of Claim 6, wherein the non-autonomous retrotransposon is an Alu element or a pseudogene.

9. The method of Claim 1 or Claim 2, wherein the gene of interest is the C4A gene; the human mobile element of interest is a retroviral insert in intron 9 of the C4A gene; and the alteration in the gene of interest is a deletion in the C4A gene.

10. The method of Claim 9, wherein the deletion in the C4A gene is C4AQ0.

11. The method of Claim 10, wherein the long range polymerase chain reaction amplification utilizes a forward primer that corresponds to a DNA sequence in the G11 gene, and a reverse primer corresponding to a DNA sequence in exon 10 of the C4A gene.

12. The method of Claim 11, wherein PCR products, if present, are approximately 5.4 kb in size.

13. The method of Claim 11, wherein the forward primer is TCTAGCTTCAGTACTTCCAGCCTGT (SEQ ID NO:1); and the reverse primer is GATGACACAAAATACCAGGATGTGA (SEQ ID NO:2).

14. The method of Claim 9, wherein the target DNA comprises a junction between intron 9 and retroviral insert in intron 9.

15. The method of Claim 14, wherein the deletion in the C4A gene is C4AQ0

16. The method of Claim 15, wherein the long range polymerase chain reaction amplification utilizes a forward primer that corresponds to a DNA sequence in the G11 gene, and a reverse primer corresponding to a DNA sequence in the junction between intron 9 and the retroviral insert in intron 9 of the C4A gene.

17. The method of Claim 16, wherein the PCR products, if present, are approximately 5.2 kb in size.

18. The method of Claim 16, wherein the forward primer is TCTAGCTTCAGTACTTCCAGCCTGT (SEQ ID NO:1); and the reverse primer is TGGTCCCCAACATGTCTGTGCATGCTG (SEQ ID NO:3).

19. A method of determining whether an individual is at risk for developing a disease associated with an alteration in a gene of interest, comprising:

a) subjecting a test sample comprising genomic DNA from the individual to long range polymerase chain reaction amplification of target DNA comprising a human mobile element, such that if the test sample comprises genomic DNA comprising an alteration, PCR products are formed, and if the test sample does not comprise genomic DNA comprising an alteration no PCR products are formed; and

b) detecting the presence or absence of the PCR products,

wherein the presence of PCR products is indicative of the presence of an alteration in the gene of interest in the test sample which correlates with a risk for developing the disease associated with the alteration in the gene of interest.

20. A method of determining whether an individual is at risk for developing a disease associated with an alteration in a gene of interest, comprising:

a) subjecting a test sample comprising genomic DNA to long range polymerase chain reaction amplification of target DNA comprising a human mobile element of interest, such that if the test sample comprises genomic DNA comprising the alteration in the gene of interest, PCR products having a first size are formed, and if the test sample does not comprise genomic DNA comprising the alteration in the gene of interest, PCR products having a second size are formed, wherein the first size and second size are detectably different; and

b) detecting the presence or absence of the PCR products,

wherein the presence of PCR products is indicative of the presence of an alteration in the gene of interest in the test sample which correlates with a risk for developing the disease associated with the alteration in the gene of interest.

21. The method of Claim 19 or Claim 20, wherein the gene is C4A gene; the disease associated with the alteration in the gene is systemic lupus erythematosus; the human mobile element of interest is a retroviral insert in

intron 9 of the C4A gene; and the alteration in the gene of interest is a deletion in the C4A gene.

22. The method of Claim 21, wherein the deletion in the C4A gene is C4AQ0.

23. A method of determining C4A deletion genotype of an individual, comprising:

a) subjecting a test sample comprising genomic DNA from the individual to long range polymerase chain reaction amplification of target DNA comprising a retroviral insert in intron 9 of the C4A gene, such that if the test sample comprises genomic DNA comprising C4AQ0, PCR products are formed, and if the test sample does not comprise genomic DNA comprising a C4AQ0, no PCR products are formed; and

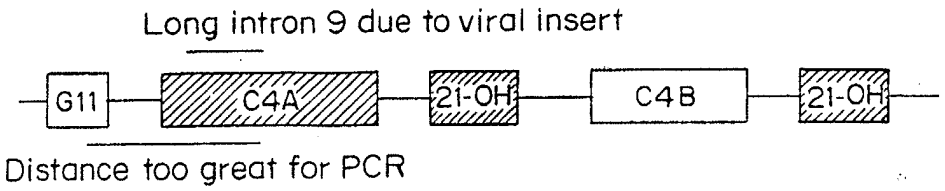
b) detecting the presence or absence of the PCR products, wherein the presence of PCR products indicates that the individual is either homozygous or heterozygous for C4AQ0, and the absence of PCR products indicates that the individual is homozygous for the absence of C4AQ0.

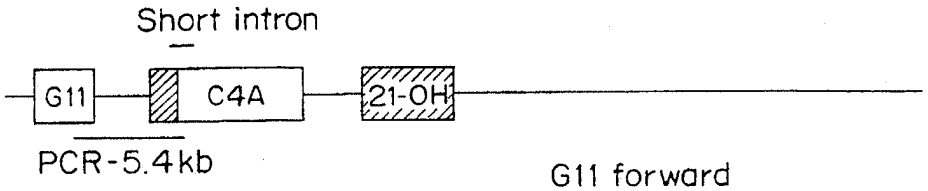
24. The method of Claim 23, further comprising:

a) subjecting a test sample comprising genomic DNA from the individual to long range polymerase chain reaction amplification of target DNA comprising a junction between intron 9 and retroviral insert in intron 9 of the C4A gene, such that if the test sample comprises genomic DNA comprising a C4AQ0, no PCR products are formed, and if the test sample does not comprise genomic DNA comprising C4AQ0, PCR products are formed; and

b) detecting the presence or absence of the PCR products,

wherein the absence of PCR products indicates that the individual is homozygous for C4AQ0, and the presence of PCR products indicates that the individual is heterozygous for C4AQ0.



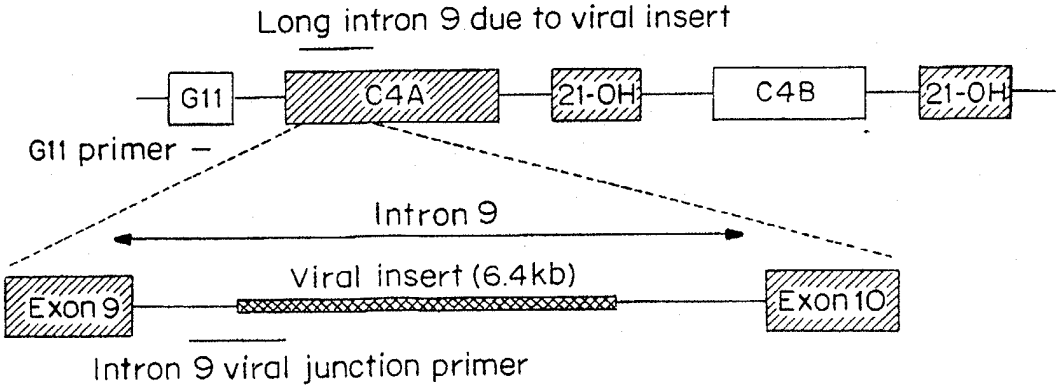


G11 forward
TCTAGCTTCAGTACTTCCAGCCTGT

Exon 10 reverse
GATGACACAAAATACCAGGATGTGA

5409bp

FIG. 1



G11 forward
TCTAGCTTCAGTACTTCCAGCCTGT

Intron 9 Viral Junction reverse
TGGTCCCCAACATGTCTGTGCATGCTG

5248bp

FIG. 2